

3

환자-치과의사 관계(PDI Patient Dentist Interaction) 평가의 오차원 분석: 일반화가능도 이론 적용

연세대학교 치과대학 치의학교육연구센터¹, 강릉원주대학교 치과대학 치과보철학교실²
¹김주아, ²조리라

ABSTRACT

Analysis of Error Source in Subjective Evaluation on Patient Dentist Interaction : Application of Generalizability Theory

Yonsei University College of Dentistry¹, Gangneung-Wonju National University College of Dentistry¹
¹Jooah Kim, ²Lee-Ra Cho

This study aims to apply the Generalizability Theory (G-theory) for estimation of reliability of evaluation scores between raters on Patient Dentist Interaction. Selecting a number of raters as multiple error sources, this study was analyzed the error sources caused by relative magnitude of error variances of interaction between the factors and proceeded with D-study based on the results of G-study for optimal determination of measurement condition. The estimated outcomes of variance component for accuracy among the Patient Dentist Interaction evaluation with G-theory showed that impact of error was the biggest influence factor in students. The second influence was the item effect, and the rater effect was relatively small. The Generalizability coefficients for case1 and case2 which were estimated through the D- study were calculated relatively low.

Key words : Generalizability Theory, Patient Dentist Interaction, Variance Component, Reliability

Corresponding Author : Jooah Kim
 Dental Education Research Center in Yonsei University College of Dentistry, 50-1, Yonsei-ro, Seodaemun-gu, Seoul, Korea
 Tel : +82-2-2228-2977, Fax : +82-2-362-8618, E-mail : kja35@yuhs.ac

I. 서론

2022년 치과대학 및 치의학전문대학원 졸업예정자들부터 치과의사 국가고시 필기시험에 실기시험을 추가로 시행하기로 하였다. 국가고시의 실기시험 도입은 치의학 교육에 커다란 영향을 미칠 것으로 예상된다. 이미 의사 국가고시 실기시험은 2010년에 도입되어, 현장에서 필요로 하는 유능한 의료인의 자질과 요건은 무엇이며, 자격을 부여하기 위한 시험 방법과 절차에 대해 많은 논의를 거쳤다고 볼 수 있다.

치과의사 국가고시 실기시험은 교과목별 시험이 아니라 교육과정에서 충분히 학습한 임상실습에 바탕을 두고 진료 과정을 반영하여 평가 가능한 문항으로 평가하는 것이다. 그 목적은 기본직무의 문제를 해결하는 자질과 함께 피험자의 종합적 사고력을 평가함으로써 치과의사의 역량을 평가하는데 있다. 따라서 임상상황 하에서 위와 같은 목적을 달성할 수 있는 유기적 연관성을 가진 실기시험 문항을 개발하기 위해 노력 중이다.

2019년에 본과 2학년으로 재학 중인 학생들이 치르게 될 치과의사 국가고시 실기시험은 진료역량을 평가하는 “가”형과 문제해결형 수기를 평가하는 “나”형으로 구분된다. 가형의 문항은 진료에 대한 준비, 상황에 대한 과정평가를 위해 각 문항에서 정확한 진단명을 끌어낼 수 있도록 문진하는지에 대한 평가, 병력에 대한 청취, 정확한 진단을 위한 검사와 진찰항목에 대한 수행여부, 진단을 바탕으로 한 예후평가와 치료계획 수립능력을 평가할 수 있어야 하며 행위가 이루어지는 과정 중에 이루어진 피험자의 행위를 바탕으로 환자-치과의사 관계(PDI, Patient Dentist Interaction)까지 평가하는 새로운 형식의 시험이다. 익숙하지 않은 방식의 평가이기는 하지만 치과대학 교육 및 임상실습을 통해 충분히 교육받은 내용을 평가하는 것이기에 평가의 의의는 충분하지만 처음 시도되는 형식이어서 사전에 여러 번의 모의평가를 통해 향후 발생할 문제를 최소화해야 한다. 이러한 평가방법은 기존의 환자-의사 관계를 평가하는

데 사용되었으나, 치과의사의 경우 전례가 없는 것으로 실기시험의 근본적인 취지에 완벽하게 부합하는 양식이라 할 것이다.

향후 치과의사 국가고시에서 평가하고자 하는 환자-치과의사 관계(PDI)는 채점자가 관찰하면서 평가하는 주관형 평가로 다양한 요인에 의해 영향을 받을 수 있다. 표준화 환자와 각 치과대학에서 선발된 평가자가 채점자로 평가할 수 있으며, 이는 과정평가이자 주관형 평가이기 때문에 평가자에 대한 체계적 교육 및 숙련도가 환자-치과의사 관계(PDI) 평가에 영향을 줄 수 있다. 임상실습에서 일반적으로 경험했던 상황인지 아닌지에 따른 피험자가 체감하는 난이도의 차이가 환자-치과의사 관계(PDI) 평가에 영향을 줄 수 있다. 채점항목도 환자-치과의사 관계(PDI) 평가에 영향을 미쳐 지나치게 복잡하거나 채점항목 간의 중첩성이 있을 때는 평가의 일관성이 담보되지 않는다. 편의(bias)가 개입되지 않은 피험자의 실제 환자-치과의사 관계(PDI)를 명확하게 판단하기 위해서는 다양한 사전시도를 통하여 어떤 요인들이 환자-치과의사 관계(PDI) 평가에 영향을 미칠 수 있는지를 면밀히 판단하고 이에 대한 대비책을 적절하게 수립해야 한다. 환자-치과의사 관계(PDI) 평가에 영향을 주는 피험자 효과, 문항효과, 채점자효과, 상호작용 효과 및 잔차 등에 대한 분석이 반드시 필요하다고 하겠다.

이에 본 연구는 강릉원주대학교 치과대학 4학년 학생들을 대상으로 두 가지 케이스를 선정해서 실시한 환자-치과의사 관계(PDI) 평가결과에 일반화가능도 이론을 적용하여 오차변량의 상대적 크기로 평가점수에 미치는 영향력과 신뢰도를 분석하였다. 연구의 구체적인 연구문제는 다음과 같다.

첫째, 환자-치과의사 관계에서 강릉원주대학교 치과대학 피험자의 평가점수에 영향을 미치는 요인들의 상대적인 영향력은 어떠한가?

둘째, 강릉원주대학교 치과대학 피험자의 환자-치과의사 관계점수의 신뢰도는 어떠한가?

II. 이론적 배경

1. 환자-치과의사관계 평가

환자-치과의사관계 평가는 2018년 10월 26일 금요일, 강릉원주대학교 치과대학 4학년 재학생 44명에게 시행하였다. 환자-치과의사 관계를 평가하기 위해 급성치수염과 치아균열에 대한 두 가지 케이스를 개발하였다. 환자-치과의사 관계 평가문항은 각 임상적 문제에 대한 대처뿐만 아니라 환자를 대하는 태도 및 의사소통능력도 함께 평가하는 문항으로 구성하였다. 급성치수염 case1은 19개의 평가문항이고 치아균열 case2는 22개 평가문항이었다.

시험 전 피험자에게 시험의 진행일정에 대한 간략한 설명과 함께 케이스 유형에 대하여 교육하였다. 환자-치과의사 관계 평가는 강릉원주대학교 치과대학에 재직하는 8인, 타 치과대학에 재직하는 4인 모두 12명의 교수가 채점자로 참여하였다. 채점자간 신뢰도를 확보하기 위해 12인의 채점자들에게 두 가지 케이스 평가에 대한 교육을 실시하였다. 채점자 교육은 case 설명, 평가문항 설명, 모의평가를 실시하고 채점결과에 대해 토의하였다.

2. 일반화가능도이론

측정점수에 영향을 주는 측정조건을 일반화가능도이론(G이론: generalizability theory)에서는 국면(facet)이라고 정의한다. 이는 분산분석에서 요인(factor) 또는 회귀분석에서 독립변수와 비교할 수 있다. G이론에서 전집(universe)은 모집단과 조금 다른 의미를 갖는다. G이론에서 전집은 단순 측정대상을 의미하는 것이 아니라 측정대상의 측정조건들에 대한 일반화 과정을 포함해서 설명하는 것이다. 전집은 한 개의 국면 또는 여러 개의 국면을 결합하여 설계하고, 측정대상(피험자 subject)은 국면으로 간주하지 않는다. 따라서 본 연구의 국면은 채점자(rater)와 채점문항(item)이었다.

한 요인의 모든 조건과 다른 요인의 모든 조건이 교차(crossed)되면 기호로 '×'를 사용하고, 한 요인이 다른 한

요인 내에 내재(nested)되면 기호를 ':'를 사용한다. 본 연구는 채점자는 피험자에 내재되었고, 피험자는 문항과 교차되었다.

G이론은 피험자를 제외한 측정국면으로 설계하여 G-연구(generalizability study)와 D-연구(decision study)로 구분한다. G-연구는 허용 가능한 관찰전집(universe of admissible observation)과 관계된 요인들의 분산성분 추정값을 산출하여, 관찰점수에 영향을 주는 오차요인의 영향력을 비교하기 위해 적용한다. D-연구는 G-연구 결과 산출된 측정점수의 분산성분 추정값을 토대로 일반화가능도 계수(generalizability coefficient)와 의존도 계수(dependability coefficient) 정보를 제공한다.

일반화가능도 계수($E\rho^2$)는 규준참조검사에서 신뢰도와 비교하여 설명할 수 있다. 예를 들어 $p \times i$ 와 같이 피험자(p)와 문항(i)이 교차하는 국면이 하나인 교차설계의 경우 Cronbach's α 와 동일한 값을 산출한다. 의존도 계수(ϕ)는 절대오차분산과 각 국면의 수와 분할점수 변화에 따른 변동의 추이를 살펴보고, 준거참조검사에서 결과의 안정성에 대한 정보를 제공할 수 있다(1. Brennan, 2001; 2. Kim & Kim, 2001; 3. Shavelson & Webb, 1991). 본 연구에서는 각 케이스의 일반화가능도 계수를 산출하여 신뢰도 비교에 목적을 두었다.

III. 연구방법

1. 분석자료

강릉원주대학교 치과대학 4학년 44명의 환자-치과의사 관계 실습시험 결과를 분석하였다. 4개의 고사장에 각 11명의 피험자(p)가 시험을 치르고 각 피험자에 대해 3인의 채점자(r)가 모든 문항(i)을 채점하는 형식으로 진행되었다. 실제 분석자료는 고사장(s)에 피험자가 내재되어 있고 채점자는 피험자에 내재되어 있는($r:p:s$)와 같은 구

Table 1. Data Structure

test site	subject (p)	rater (r)	item(i)			
			i_1	i_2	...	i_n
1	p_1	r_1				
		r_2				
		r_3				
	...					
	p_{11}	r_1				
		r_2				
		r_3				
2	p_{12}	r_4				
		r_5				
		r_6				
	...					
	p_{22}	r_4				
		r_5				
		r_6				
3	p_{23}	r_7				
		r_8				
		r_9				
	...					
	p_{33}	r_7				
		r_8				
		r_9				
4	p_{34}	r_{10}				
		r_{11}				
		r_{12}				
	...					
	p_{44}	r_{10}				
		r_{11}				
		r_{12}				

조이나, 측정대상인 피험자는 내재될 수 없는 성분이므로 고사장은 국면으로 고려하지 않았다. Table 1은 분석자료 구조를 표로 나타낸 것이다.

2. 분석방법

1) G-연구 설계

채점자(r)는 시험을 치르는 피험자(p)에게 내재되어 있고, 각 피험자는 모든 문항(i)을 평가 받는 구조이다. 일 변화가능도 이론의 2국면 부분내재(two-facet nested design)인 $(r:p) \times i$ 설계를 적용하였다.

평가 상황에서 피험자에게 부여한 점수는 식(1)과 같이 차례대로 총 평균, 피험자 효과, 문항 효과, 피험자에 내재된 채점자 효과, 피험자와 문항의 상호작용 효과 그리고 잔차 효과로 분리할 수 있으며 각 효과를 도식화한 것은 Fig. 1에 제시하였다.

검사점수의 분산은 식(2)와 같이 5개의 독립적인 분산 성분의 합으로 설명할 수 있다. $\sigma^2(p)$ 는 피험자 간 차이, $\sigma^2(i)$ 는 문항 간 난이도 차이, $\sigma^2(r:p)$ 는 피험자에 내재된 채점자간 차이, $\sigma^2(pi)$ 는 문항에 따른 피험자간 차이, $\sigma^2(ir:p)$ 는 피험자에 내재된 채점자와 문항의 상호작용을 포함한 잔차에 의한 차이를 나타낸다.

식(1)

$$\begin{aligned} X_{pri} = & \mu \\ & + (\mu_p - \mu) \\ & + (\mu_{r:p} - \mu_p) \\ & + (\mu_i - \mu) \\ & + (\mu_{pi} - \mu_p - \mu_i + \mu) \\ & + (\mu_{pri} - \mu_{pi} - \mu_{r:p} + \mu_p) \end{aligned}$$

총 평균
피험자 효과
피험자에 내재된 채점자 효과
문항효과
피험자와 문항의 상호작용효과
잔차

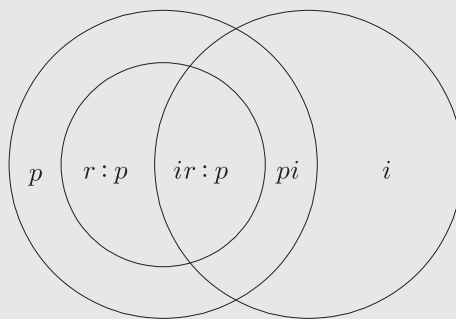


Fig 1. G-study for $(r:p) \times i$ design

식(2)

$$\sigma^2(x) = \sigma^2(p) + \sigma^2(r:p) + \sigma^2(i) + \sigma^2(pi) + \sigma^2(ir:p)$$

2) D-연구 설계

D-연구 설계는 G-연구설계와 동일하게 적용하였다. 전집 분산성분($\sigma^2(\tau)$), 상대오차 분산성분($\sigma^2(\delta)$), 절대오차 분산성분($\sigma^2(\Delta)$)은 식(3)–식(5)에서 얻을 수 있다.

$$\sigma^2(\tau) = \sigma^2(p) \quad \text{식(3)}$$

$$\sigma^2(\delta) = \frac{\sigma^2(r:p)}{n_r} + \frac{\sigma^2(pi)}{n_i} + \frac{\sigma^2(ir:p)}{n_i n_r} \quad \text{식(4)}$$

$$\sigma^2(\Delta) = \frac{\sigma^2(r:p)}{n_r} + \frac{\sigma^2(i)}{n_i} + \frac{\sigma^2(pi)}{n_i} + \frac{\sigma^2(ir:p)}{n_i n_r} \quad \text{식(5)}$$

규준참조검사에 적용되는 신뢰도인 일반화가능도 계수

($E\rho^2$), 그리고 준거참조검사에 적용되는 의존도계수(Φ)는 식(6)–식(7)로 산출하였다.

$$E\rho^2 = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\delta)} \quad \text{식(6)}$$

$$\Phi = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\Delta)} \quad \text{식(7)}$$

IV. 연구결과

1. G-연구를 통한 검사점수에 영향을 미치는 요인들의 상대적 영향력 탐색

환자-치과의사 관계를 평가한 점수에 영향을 미치는 요인들의 상대적인 영향력을 탐색하기 위해 채점자와 문항요인을 무선효과로 가정하는 $(r:p) \times i$ 설계를 적용하여 분

Table 2. Variance components for $(r:p) \times i$ design

case	Source of Variance	Degree of freedom	sum of squares	Mean Squares	Estimated variance component	Percentage of total variance
case1 (Acute pulpitis)	<i>p</i>	43	32.5128	0.7561	0.0095	59.66%
	<i>r:p</i>	88	6.9737	0.0793	0.0008	5.05%
	<i>i</i>	18	119.0927	6.6163	0.0026	16.22%
	<i>pi</i>	774	1596.6667	133.0213	0.0024	15.34%
	<i>ri:p</i>	1584	1656.5000	52.8597	0.0006	3.72%
	total	2507	3411.7458		0.0158	100.0%
case2 (Crack tooth)	<i>p</i>	43	34.8409	0.8103	0.0078	48.08%
	<i>r:p</i>	88	7.1515	0.0813	0.0004	2.54%
	<i>i</i>	21	241.9205	11.5200	0.0039	23.98%
	<i>pi</i>	903	2025.5000	244.5720	0.0033	20.32%
	<i>ri:p</i>	1848	2133.0000	100.3485	0.0008	5.08%
	total	2,507	4442.4129		0.0161	100.00%

석한 G-연구 결과는 Table 2와 같다.

case1의 피험자 효과가 59.66%로 가장 큰 영향력을 미치는 것으로 나타났다. 이는 피험자들의 환자-치과의사 관계 능력의 차이가 점수에 가장 크게 영향을 미치는 것으로 나타났다. 두 번째는 평가문항의 난이도 효과는 16.22%로 나타났고 문항에 따른 피험자 효과는 15.34%로 나타났다. 반면 채점자 효과는 5.05%로 비교적 적은 부분을 차지하였다.

case 2의 피험자 효과가 48.08%로 가장 큰 영향력을 미치는 것으로 나타났다. case1과 비슷하게 피험자들의 환자-치과의사 관계 능력의 차이가 점수에 가장 크게 영향을 미치는 것으로 나타났다. 두 번째는 평가문항의 난이도 효과로 23.98%로 나타났고 문항에 따른 피험자 효과는 20.32%로 나타났다. case1과 비교하였을 때 평가문항 효과가 높게 나타난 이유에 대해 case2가 더 어려운 문항으로 피험자들이 인식한 것으로 분석할 수 있다. 반면 채점자 효과는 2.54%로 case1보다도 적은 부분을 차지하였다.

2. 신뢰도 산출을 위한 D-연구 결과

G-연구와 동일한 $(R: p) \times I$ 설계의 D-연구를 적용했을 때의 오차분산성분과 신뢰도를 분석한 결과를 Table 3에 제시하였다. 환자-치과의사관계 점수를 표준참조검사로 사용할 때의 오차분산인 상대오차분산이 준거참조검사로 사용할 때의 오차분산인 절대오차분산보다 작게 나타났다.

이는 D-연구 분석에서 흔히 나타날 수 있는 결과로, 절대 오차분산에는 상대오차분산 외에 문항 효과가 합쳐져 있기 때문이다. 환자-치과의사 관계 점수를 표준참조검사로 사용할 때의 신뢰도인 일반화가능도계수가 준거참조검사로 사용할 때의 신뢰도인 의존도계수보다 더 크게 나타났다. 역시 D-연구 분석에서 흔히 나타날 수 있는 결과로, 절대 오차분산이 상대오차분산보다 더 크게 나타났기 때문이다.

V. 고찰

본 연구는 강릉원주대학교 치과대학 4학년 44명을 대상으로 환자-치과의사관계 평가를 실시한 결과에 일반화가능도 분석을 수행하였다. 이를 바탕으로 치과의사 국가고시 실기시험을 앞둔 졸업생의 환자-치과의사관계 평가 점수에 영향을 미치는 요인들의 상대적인 영향력을 탐색하고, 평가점수에 영향을 미치는 요인의 분산성분에 차이가 있는지, 그리고 신뢰도에 차이가 있는지를 분석하였다.

첫째, 환자-치과의사관계 평가점수에 영향을 미치는 요인들의 상대적인 영향력을 탐색하기 위해 G-연구를 분석한 결과, 피험자 효과가 가장 크게 나타났다. 이는 환자-치과의사 관계 평가점수가 학생의 환자-치과의사 관계 역량에 따른 수준 차이를 반영하고 있음을 의미한다. 이는 PDI 평가가 다른 요인들보다 수험생의 역량에 따른 결과를 보

Table 3. Summary of D-study results

case	# of raters	# of items	Relative error variance	Absolute error variance	Generalizability coefficient	Dependability coefficient
case1 (Acute pulpitis)	3	19	0.0038	0.0064	0.7120	0.5965
case2 (Crack tooth)	3	22	0.0045	0.0084	0.6325	0.4808

인다는 극히 중요한 결론을 도출해냄으로써 본래 의도하였던 출제 및 평가의도와 일치한다는 점을 확인할 수 있었다. 이에 각 교육기관에서는 평소의 교육을 통해 올바른 환자-치과의사 관계가 정립될 수 있도록 부단한 노력을 기울여야 할 것이다. 두 번째로 큰 영향력은 문항 효과로 나타났다. 특히 어려운 문항에 대해서는 더 큰 영향력을 미치고 있어 문항 출제에 주의가 필요함을 알 수 있었다. 채점자 효과는 상대적으로 작게 나타났다. 환자-치과의사관계는 채점자가 피험자의 실기를 관찰하면서 평가하게 되어 채점자의 주관성이 우려되었으나 채점자 3인을 배치함에 따라 채점자 효과는 크지 않았다. 반면에 본 연구 설계에서 고려하지 못한 국면들의 필요성을 간접적으로 설명하는 잔차 효과는 작게 나타났다. 이처럼 일반화가능도 분석은 환자-치과의사관계 평가점수 차이가 피험자들의 환자-치과의사관계 역량 수준 차이 때문인지, 문항의 난이도 차이 때문인지, 채점자 특성 때문인지 등을 포함하여 본 연구에서 고려하지 못한 추가적인 요인들이 필요한지를 밝히는데 활용할 수 있다.

둘째, 신뢰도를 산출하기 위해 D-연구를 분석한 결과 case1이 case 2보다 신뢰도가 높게 산출되었다. 신뢰도의 적정 수준에 대한 기준에 대해 Brennan(2001)과 Fyans(1983)는 의존도계수의 경우 0.70, Norcini(1999)는 검사 종류에 상관없이 0.80으로 제시하였다.^{1) 4) 5)} 그러나 본 연구결과에서는 이에 못 미치는 수치가 산출되었다. 이는 관찰에 의한 평가인 점과 평가문항 수가 19개와 22개인 것을 고려하여 해석할 필요가 있다.

이러한 분석결과를 토대로 G이론의 적용에 대한 시사점을 제시하면 첫째, 환자-치과의사관계 평가점수에 G-연구를 수행함으로써 구체적으로 평가점수에 영향을 미치는 피험자 역량차이, 문항, 채점자 등과 같은 오차요인들의 상대적인 영향력을 파악할 수 있다. 따라서 본 연구는 G이론을 환자-치과의사관계 평가점수 분석에 적용함으로써 G이론의 경험적 타당성을 높였다는데서 학문적인 의의가 있다. 둘째, 환자-치과의사관계 평가점수에 D-연구 분석을 수행함으로써 주관형 평가에서 신뢰도를 산출하여 몇 개의 문항, 몇 명의 채점자가 필요한지를 결정하는데 기여할 수 있다. 본 연구에서는 평가점수의 결과 분석과 검사의 수정 및 보완의 근거 모색에 G이론의 방법론을 적용함으로써 보다 정확하고 체계적인 분석을 시도하였다는 점에서 방법론적 의의가 있다.

마지막으로 본 연구의 제언은 다음과 같다. 첫째, 본 연구는 환자-치과의사관계 평가점수를 일반화가능도 분석하여 문항과 채점자 요인을 분리하였다. 연구에서 고려하지 못한 다른 오차요인들을 설계에 반영하여 일반화가능도 분석을 수행할 필요가 있다. 둘째, 본 연구에서는 강릉원주대학교 치과대학 4학년 학생의 환자-치과의사관계 평가점수가 일관성 있게 적용될 수 있는지의 여부에 대한 검정을 한 것이므로, 환자-치과의사관계 평가점수 자체의 신뢰성을 담보하는 것은 아니다. 따라서 환자-치과의사관계 평가를 받아야 하는 전체 피험자를 대표할 수 있는 연구대상을 표집하여 일반화가능도 분석을 수행할 필요가 있다.

참 고 문 헌

1. Brennan, R. L. (2001). Generalizability Theory. New York: Springer.
2. Kim, S. S., & Kim, Y. B. (2001). Generalizability Theory. Seoul: Kyoyookgwahaksa.
3. Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, 30(3), 215-232.
4. Fyans, L. J. (1983). Generalizability theory: Inferences and practical applications. Jossey-Bass Inc Pub.
5. Norcini, J. J. (1999). Measurement issues in the use of simulation for testing professionals: Test development, test scoring, standard setting. *Innovative simulations for assessing professional competence*. Chicago, IL: University of Illinois.